

93201Q



932012

NEW ZEALAND QUALIFICATIONS AUTHORITY  
MANA TOHU MĀTAURANGA O AOTEAROA

## Scholarship 2009 Statistics and Modelling

2.00 pm Saturday 14 November 2009

Time allowed: Three hours

Total marks: 48

### QUESTION BOOKLET

Pull out the Formulae and Tables Booklet S–STATF from the centre of this booklet.

There are SIX questions in this booklet. Answer ALL questions.

Write your answers in the Answer Booklet 93201A.

Show all working. Start your answer to each question on a new page. Carefully number each question.

There is a grid provided on page 24 of the Answer Booklet for answering Question One (a), another grid on page 25 for answering Question Two (a)(i), and a further grid on page 26 for answering Question Six (a).

Check that this booklet has pages 2–8 in the correct order and that none of these pages is blank.

**YOU MAY KEEP THIS BOOKLET AT THE END OF THE EXAMINATION.**

You have three hours to complete this examination.

The theme of this paper is real estate in a large city. Questions One, Two and Three refer to a real estate database that contains data on approximately 80 000 homes in two suburbs, North and South, of which approximately 32 000 are in North and approximately 48 000 are in South.

### QUESTION ONE (8 marks)

A sample of 200 homes was selected from the real estate database described above. The number of homes selected from each suburb is proportional to the number of homes in North and South respectively in the population (which is known as proportional allocation). Summary statistics for the home valuations, in thousands of dollars (\$000), are shown in Table 1 below.

Suburb	North	South
Sample Size	80	120
Mean	421	613
Standard Deviation	35	50
Lowest Value	361	431
Lower Quartile	390	500
Median	418	511
Upper Quartile	450	600
Highest two values	480 and 485	725 and 775

**Table 1**

- Draw a graph that enables the valuations from the two suburbs to be compared, and discuss what your graph shows about the valuations. A grid is provided on page 24 of the Answer Booklet to help you answer this question.
- Describe how the sample of 200 homes could have been selected.
- Construct a 95% confidence interval for the difference in the population mean valuations between the two suburbs.
- Make a conclusion from the interval calculated in (c).
- Estimate the mean valuation for all the homes in the database.
- Calculate the number of further valuations that should be sampled from each suburb (maintaining proportional allocation), using an estimate of \$97 000 for the population standard deviation, so that at the 99% level of confidence the population mean valuation is estimated to within \$9 000.

**QUESTION TWO** (8 marks)

- (a) A model is to be created to describe the selling price of a three-bedroom home in terms of its distance from the centre of the city. Let  $S$  be the selling price, in thousands of dollars (\$000), and  $D$  be the distance, in kilometres, from the centre of the city.

The following two non-linear models will be considered:  $S = aD^n$  and  $S = be^{kD}$  where  $a$ ,  $b$ ,  $k$  and  $n$  are constants.

Data for eight homes are shown in Table 2 below.

Distance $D$ (km)	1	5	11	13	14	17	21	25
Selling price $S$ (\$000)	575	494	365	343	331	285	240	207

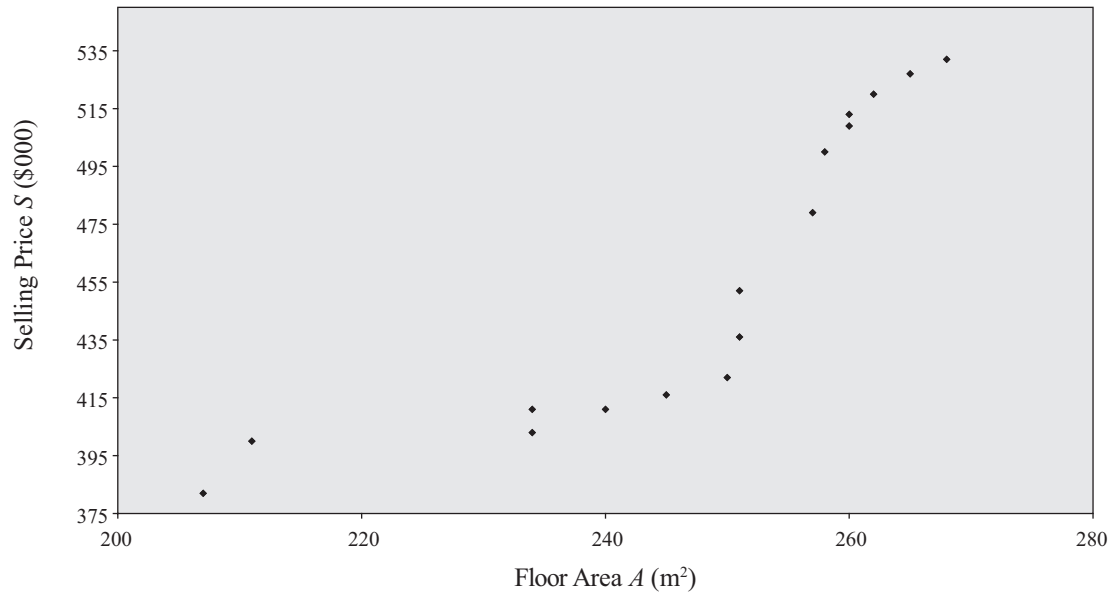
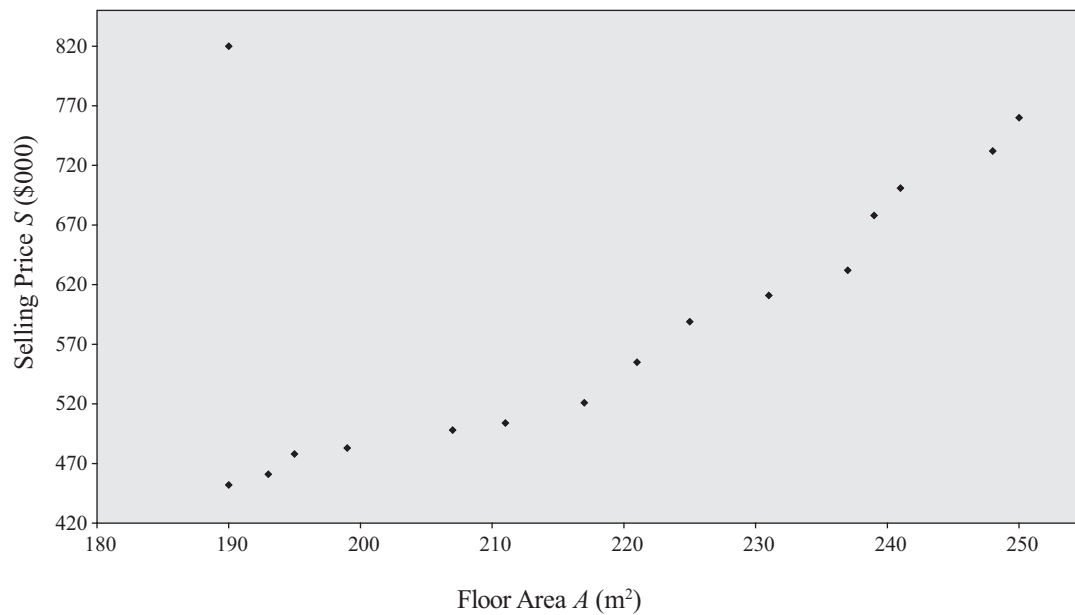
**Table 2**

- (i) Choose the more appropriate model and obtain estimates for the constants in your chosen model. Justify your choice of model. A grid is provided on page 25 of the Answer Booklet to help you answer this question.
- (ii) Predict the selling price of a three-bedroom home that is 37 km from the centre of the city, and comment on the validity of this prediction.
- (iii) Comment on the limitations of using these non-linear models to predict selling prices of three-bedroom homes.
- (b) Information in the real estate database revealed that:
- 40% of homes with an outdoor barbeque area also have a pool
  - 80% of homes with a pool also have an outdoor barbeque area
  - 5% of homes had neither a pool nor an outdoor barbeque area.

Calculate the percentage of homes that have both a pool and an outdoor barbeque area.

**QUESTION THREE (8 marks)**

A sample of 32 homes was selected from the real estate database, 16 from North and 16 from South. The selling price,  $S$ , in thousands of dollars (\$000), and the floor area,  $A$ , in square metres ( $\text{m}^2$ ), were recorded for each home selected. A scatter plot for each suburb is shown below in Figures 1 and 2.

**Figure 1: North Suburb****Figure 2: South Suburb**

Using regression analysis for each suburb, two different lines or curves were fitted to the data. The regression equations are shown below.

**North**

Equation 1:  $S = 2.5475A - 172.33$

Equation 2:  $S = 0.071A^2 - 31.214A + 3813$

**South**

Equation 3:  $S = 4.9163A - 505.9$

Equation 4:  $S = 3.2653A - 120.88$

- (a) Write a short paragraph to describe the relationship between selling price and floor area for each suburb.
- (b) For each suburb, choose the equation which is the better fit to the data and justify your choice. From your choices, obtain two predictions, one from each suburb, of the selling price for a home with a floor area of 255 m<sup>2</sup>.

Comment on the validity of each prediction.

- (c) It was suggested that the factors “age of home” and “location of home” could have an influence on the selling price and that they should be part of a more extensive analysis.

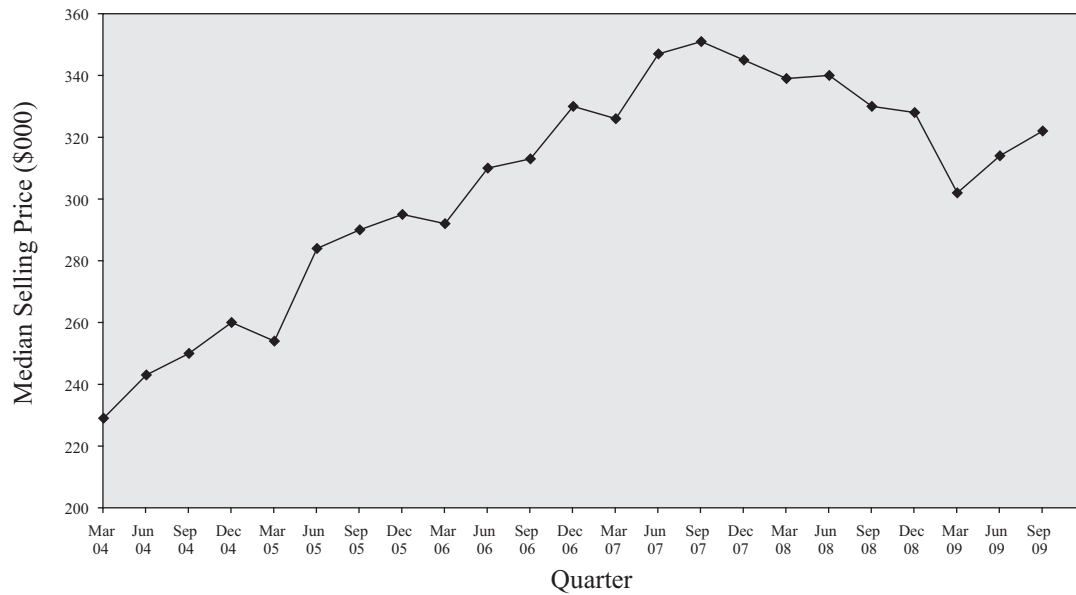
Describe the sort of relationship you would expect each of these factors to have with the selling price.

**QUESTION FOUR (8 marks)**

In an investigation of home prices, the median home selling price, in thousands of dollars (\$000), was obtained for each quarter from March 2004 to September 2009. Centred moving means (rounded to the nearest thousand dollars) were calculated. The data are shown in Table 3 below, along with an inflation index (Base: June quarter 2005 = 1000). The median home selling prices are displayed in Table 3 and a graph of median home selling prices by quarter is displayed in Figure 3.

<b>Quarter</b>	<b>Median Home Selling Price (\$000)</b>	<b>Centred Moving Mean (\$000)</b>	<b>Inflation Index</b>
Mar 04	229		
Jun 04	243		924
Sep 04	250	249	
Dec 04	260	257	
Mar 05	254	267	
Jun 05	284	276	1000
Sep 05	290	286	
Dec 05	295	294	
Mar 06	292	300	
Jun 06	310	307	1047
Sep 06	313	316	
Dec 06	330	324	
Mar 07	326	334	
Jun 07	347	340	1208
Sep 07	351	344	
Dec 07	345	345	
Mar 08	339	341	
Jun 08	340	336	1224
Sep 08	330	330	
Dec 08	328	322	
Mar 09	302	318	
Jun 09	314		1249
Sep 09	322		

**Table 3**

**Figure 3: Median Home Selling Prices, March 2004 to September 2009**

Several lines (given below) were fitted to the data and the following equations were obtained, where  $y$  is the median selling price (\$000) and  $x$  is the number of quarters since December 2003 (ie,  $x = 1$  for the March 2004 quarter,  $x = 2$  for the June 2004 quarter, etc).

Equation 1:  $y = 4.27x + 253$  with  $R^2 = 0.64$  fitted to the median sale prices.

Equation 2:  $y = 4.63x + 254$  with  $R^2 = 0.72$  fitted to the centred moving means.

Equation 3:  $y = -5.66x + 437$  with  $R^2 = 0.99$  fitted to the centred moving means for the period December 2007 quarter to March 2009 quarter.

- Write a short paragraph to describe median home selling prices from 2004 to 2009.
- Calculate a forecast for the median home selling price for the March quarter 2010. Show full working and justify your method. Discuss the validity of your forecast.
- Develop an index number series for the median home selling price that gives a value for the June quarter each year with the same base as the inflation index. Comment on the change in median home selling prices compared with inflation in the June quarters over the years 2004 to 2009.

**Questions Five and Six  
are on the following page.**

**QUESTION FIVE (8 marks)**

The floor area of new homes, in  $\text{m}^2$ , is modelled by a normal distribution with mean 245 and standard deviation 34. Overall, 30% of new homes are built with an outdoor barbeque area, and 2% of new homes require alterations after construction.

- (a) A random sample of 50 new homes is to be taken.
- Find the probability that no more than 20 homes in the sample will have an outdoor barbeque area.
  - Find the probability that the mean floor area of these 50 homes will be at least  $250 \text{ m}^2$ .
- (b) Another random sample of  $n$  new homes is to be taken.

Find the value of  $n$  so that the probability that the sample contains no homes requiring alterations is approximately 0.6.

**QUESTION SIX (8 marks)**

A proposed housing development will have two types of home, single detached homes and duplex units (a block of two homes). Two of the resources available are six hectares ( $60\,000 \text{ m}^2$ ) of land for housing and a building budget of \$7 million.

A single detached home requires  $600 \text{ m}^2$  of land and costs \$65 000 to build. A duplex unit requires  $800 \text{ m}^2$  of land and costs \$100 000 to build.

The profit from the sale of a duplex unit is 50% more than that from a single detached home.

Let  $x$  be the number of single detached homes and  $y$  be the number of duplex units. Assume that all homes built are sold.

- Find the number of each type of home that should be built to maximise the profit.  
A grid is provided on page 26 of the Answer Booklet to help you answer this question.
- A new council regulation requires that there must be at least three single detached homes for every two duplex units in the new development.

Calculate the percentage reduction in maximum profit resulting from this regulation.

- From previous similar developments, it has been shown that, for social reasons, it is desirable that the relationship between the numbers of the two types of homes built is approximately modelled by  $y = 35e^{-0.014x}$ .

Taking this relationship into account and the constraint in (b), calculate the amount of unused land and/or unused budget when  $x$  and  $y$  are such that the profit is maximised.